INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/3552

73-26,314

FEIR, Betty Jean Taub, 1943-
THE ANOVA F-TEST VERSUS THE KRUSKAL-WALLIS TEST:
A ROBUSTNESS STUDY.

The University of Oklahoma, Ph.D., 1973
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© 1973

BETTY JEAN TAUB FEIR

ALL RIGHTS RESERVED

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE ANOVA F-TEST VERSUS THE KRUSKAL-WALLIS TEST: A ROBUSTNESS STUDY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

BETTY TAUB FEIR

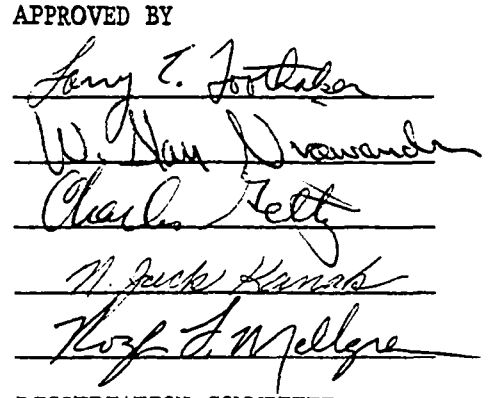
Norman, Oklahoma

1973

THE ANOVA F-TEST VERSUS THE KRUSKAL-WALLIS TEST:

A ROBUSTNESS STUDY

APPROVED BY



DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

My Committee,

Larry E. Toothaker, Chairman

W. Alan Nicewander

N. Jack Kanak

Charles Gettys

Roger Mellgren

My Parents,

Mr. & Mrs. Sam Taub

My Typist,

Marla Jardon

Very simply, with heartfelt appreciation, "I THANK YOU".

DEDICATION

This dissertation is dually dedicated to:

Pouchie, my grandmother, whose wisdom, thirst for knowledge, faith, and love encouraged me throughout my educational endeavor--

and to

Tammy, my little daughter, whose love and devotion never faltered as she waited the many hours while I studied.

TABLE OF CONTENTS

Page

Manuscript to be submitted for publication

INTRODUCTION 1

PROCEDURE. 7

RESULTS. 10

SUMMARY AND CONCLUSIONS. 19

REFERENCES. 24

APPENDIX A. . . Tables. 28

THE ANOVA F-TEST VERSUS THE KRUSKAL-WALLIS TEST: A ROBUSTNESS STUDY

ABSTRACT

The Kruskal-Wallis test and the ANOVA F-test were empirically compared in terms of probability of a Type I error and power under various combinations of assumption violations, patterns of mean differences, patterns of variance inequality, and patterns of sample size inequality. The Kruskal-Wallis test was found to be competitive to the ANOVA F-test in terms of alpha but not for power. Power of the Kruskal-Wallis test was grossly affected for nonstepwise mean differences when sample sizes and variances were negatively related and when small levels of significance were utilized. The ANOVA F-test, however, was found to be generally robust for the type of mean differences specified.

THE ANOVA F-TEST VERSUS THE KRUSKAL-WALLIS TEST: A ROBUSTNESS STUDY

INTRODUCTION

A common problem in applied research is to decide whether or not sample differences in central tendency reflect true differences in parent populations. It is appropriate to use the one-way fixed effects ANOVA F-test for the K-sample case ($K \geq 2$) if assumptions of normality, homogeneity of variances, and independence of errors are met. When either normality or equality of variances is doubtful, the use of nonparametric statistical procedures is often recommended (Bradley, 1968).

At this point, some clarification of the terms used in the nonparametric area is necessary. The term "nonparametric" is a description of the hypothesis and the term "distribution-free" is a description of the method used to test the hypothesis. Distribution-free procedures make no assumptions concerning the shape of the underlying distribution other than the assumption of continuity. Also, most distribution-free techniques were designed to test nonparametric hypotheses. That is, these techniques were generally designed to test hypotheses other than those concerning parameters, such as equality of populations. However, many distribution-free methods are commonly and most frequently used to test parametric hypotheses, primarily due to the fact that some of these methods are sensitive to not just any differences in the populations, but to differences in the parameters.

Thus, a parametric problem may or may not be tested by a method which is distribution-free, and, vice versa. However, the term "nonparametric" through usage has come to refer to procedures which may be used to test hypotheses concerning parameters, but which make no distributional assumptions other than the assumption of continuity (Kendall and Stuart, 1961). Throughout this paper, "nonparametric" and "distribution-free" will be used interchangeably in a colloquial rather than literal sense.

The two-sample Mann-Whitney U test is a frequently used distribution-free analogue to the Student's t-test, and in a one-way ANOVA situation, the distribution-free analogue to the F-test is the Kruskal-Wallis rank test (Kruskal, 1952). Both of these distribution-free methods have been commonly used to test hypotheses about means. In fact, they are generally sensitive to differences in location, but specifically sensitive to differences in medians. If the populations are symmetric, then the means and the medians are the same. For location differences, the Kruskal-Wallis test as compared to the F-test has an asymptotic relative efficiency of .95 for the normal case, 1.00 for the uniform case, and a lower bound of .864 (Bradley, 1968). Thus, not more than 13.6% efficiency can be lost using the Kruskal-Wallis test rather than the F-test, and, if the distribution is normal, the asymptotic loss of efficiency is only about 5%. Keeping in mind that asymptotic relative efficiencies are computed for unrealistically large sample sizes with minuscule differences in measures of location, it would seem profitable to investigate the efficiency of the Kruskal-Wallis test as compared to the F-test for realistic sample sizes and realistic differences in location (Feir and Toothaker, 1972).

Research has demonstrated, however, that in certain cases the ANOVA F-test as well as other parametric procedures can be said to be "robust" or insensitive to assumption violation (Box and Andersen, 1955). Robustness creates a dilemma for the researcher: should he or should he not utilize nonparametric procedures in lieu of parametric procedures? Many studies have investigated the effect of assumption violation on parametric procedures, but very few have directly compared parametric to nonparametric techniques under assumption violation.

If the assumption of independence of errors in ANOVA is violated, the consequences can be serious. Positive correlations will generally yield a liberal test while negative correlations will produce a conservative test (Cochran, 1947; Scheffé, 1959).

When heterogeneous variances exist, the one-way ANOVA F-test has been shown to be relatively robust if sample sizes are equal (Hsu, 1938; Box, 1954; Scheffé, 1959; Pratt, 1964). For normal distributions with unequal sample sizes and unequal variances, the actual alpha has been found to be smaller than set alpha (a conservative test) when there is a positive relationship between sample sizes and variances. Conversely, with negatively related sample sizes and variances, actual alpha has been found to be larger than set alpha, thus engendering a test that is too liberal.

It has been reported that Student's t-test and the ANOVA F-test are insensitive to the singular violation of normality even for small samples of equal sample size (Rider, 1929; Pearson, 1931; Scheffé, 1959; David and Johnson, 1951; Norton, 1952; Lindquist, 1953; Boneau, 1960). However, one study illustrated sensitivity of the t- and F-tests to extreme non-normality in the extreme tail regions, that is, $\alpha \leq .01$ (Bradley, 1963, 1966). Glass (1972), in retort to Bradley's findings, stated:

> "Significance at the .01 level should be sufficiently convincing that appeal to untrustworthy smaller alphas is unnecessary. Arguments that significance at smaller alpha levels should be reported to emphasize strength of effects are unconvincing. Effects should be measured on the metric of the dependent variable and should be bracketed by appropriate confidence intervals; they should not be measured with alpha levels."

Other studies have investigated the dual effect of non-normality and heterogeneous variances on alpha and found empirical alpha to be somewhat deviant from nominal alpha (Norton, 1952; Boneau, 1960; Lunney, 1970). That the degree of deviance was found to vary may be due to a variety of conditions such as inequality of sample sizes, level of significance, total sample size, type of non-normality, and so on.

Some studies have not only investigated probability of Type I error for ANOVA, they have also investigated power (Pearson, 1929; Horsnell, 1953; Srivastava, 1959; Scheffé, 1959; Games and Lucas, 1966; Lunney, 1968; Donaldson, 1968). These studies found varying power for various combinations of assumption violation. For moderate departures from normality, little difference has been found between normal theory power calculations of ANOVA based on the non-centrality parameter and the empirical power values for non-normal populations. For populations which were extremely skewed or markedly leptokurtic, there was a significant positive effect on power (Games and Lucas, 1966). Donaldson (1968) found that, generally, power of the F-test for non-normal distributions was higher than that for normal distributions except for very large values of power.

Investigating the power of the simple t-test, Pearson (1929) found that for positively skewed populations the actual power exceeds the nominal power (H_o: $\mu=a$) if $\mu-a$ is positive. For negatively skewed populations,

Pearson found that, if μ-a is negative, the actual power is less than the nominal power. The relationship Pearson discerned was not substantiated by Strivastava's 1959 non-empirical investigations into the effect of skewness and kurtosis on both the simple t-test and the ANOVA F-test. In 1962, Boneau empirically found that platykurtic distributions produced power slightly below that of normal distribution theory and highly leptokurtic distributions produced power above that of normal distribution theory. Equal sample sizes for the two sample t-test diminished the effect of non-normality on power, and the effect of kurtosis on power was seen to be greater than the effect of skewness (Pearson, 1929).

For unequal variances, theoretical power has been determined through use of the average variance in the non-centrality parameter. This procedure has been questioned, but Donaldson's study (1968) showed for normal distributions a close agreement between actual power as calculated by Horsnell (1953) and theoretical power calculated using the average variance. Donaldson also investigated the combined effect of non-normality and heterogeneous variances, but there remains some question concerning the validity of the procedure used in calculating a theoretical power for non-normal distributions with unequal variances. Further investigation into this area is indicated.

In one of the few studies which directly compared paramtetric procedures to their nonparametric counterparts, Pratt (1964) compared Student's t-test with three nonparametric procedures (the median test, the Mann-Whitney U test and the normal scores test) for large samples, investigating both normal and symmetric non-normal distributions. Although

distribution-free tests do not require normality, Pratt found that they are, like parametric tests, sensitive to unequal population variances. For example, when nominal alpha is .05 for a variety of unequal variances, exact alpha for all four tests exhibited deviations from nominal alpha. Other studies (Toothaker, 1972, 1973) empirically compared Student's t-test, the nonparametric permutation t-test and the Mann-Whitney U test in terms of probability of a Type I error and power. Under assumption violation, the permutation t-test was found to provide empirical alphas and power which were generally close to those of Student's t-test. With unequal variances, the probabilities of the permutation t-test deviated in the same manner as those of Student's t-test and the Mann-Whitney U test. For a skewed population with unequal sample sizes and unequal variances, reversals were discovered. Reversed were the normal theory conservativism of alpha for positive relationships between sample sizes and variances, and the normal theory liberalism for negative relationships. Bradley (1968), investigating only the Student's t-test, noted the same reversals and demonstrated them to be sample size dependent.

Nonparametric procedures for the K-sample case have not been extensively studied as has ANOVA. Nevertheless, literature often recommends these distribution-free techniques as being powerful alternatives when parametric assumptions are violated (Terry, 1952; Hodges and Lehman, 1961; Klotz, 1963; Van der Laan, 1964; Puri, 1964; McSweeney, 1967; Bradley, 1968; Penfield and McSweeney, 1968; McSweeney and Penfield, 1969). McSweeney and Penfield (1969) cited theoretical and empirical evidence in support of using the normal scores test in preference to the Kruskal-Wallis test. This support was not substantiated by Feir and Toothaker (1972) who compared

the ANOVA F-test, the Kruskal-Wallis test, and the normal scores test in terms of empirical alpha and empirical power with samples from the normal distribution, and two exponential distributions. When testing hypotheses about means, empirical evidence supported the use of the ANOVA F-test over the nonparametric methods even under violation of assumptions. The Kruskal-Wallis test was found to be competitive to the F-test when testing hypotheses about medians; but in the cases investigated, the normal scores test was not consistently better than the F-test or the Kruskal-Wallis test.

Despite the multiplicity of studies which have been done on the ANOVA F-test under assumption violation, few direct comparisons of the F-test to its nonparametric counterpart, the Kruskal-Wallis test, have been made. Since under assumption violation the F-test appears to be robust and the Kruskal-Wallis test appears to be powerful, it would seem profitable to compare the two tests under extensive combinations of parametric assumption violations. The present research empirically investigates the effect of non-normality coupled with a variety of unequal variances, and several types of mean differences for the F-test and the Kruskal-Wallis test for five levels of significance. Sample size per treatment level as well as total sample size are manipulated along with the assumption violations.

PROCEDURE

The basic numbers upon which this Monte Carlo study was based were obtained via a computer using a pseudo-random number generator. Depending upon the assumption violation, the numbers were selected from either a normal distribution or from an exponential distribution scaled to have

equal medians of zero value under the null hypothesis. The random deviates were allocated to four treatment levels which comprised a one-way fixed effects analysis of variance situation.

The observations from the normal distribution were derived by a technique developed by Box and Muller (1958), which generates pseudo-random variables distributed N(0,1). For the null situation, the means of the four treatment levels were zero. The non-null situation was established by defining values of μ_j (see Table 1), j = 1,2,3,4; such that the power for the ANOVA F-test would be about .86 for large mean differences and about .60 for smaller mean differences for the equal variances condition for the normal distribution. For each of the theoretical power values, both stepwise (1:2:3:4) and nonstepwise (1:4:4:4) differences were determined. Specification of the μ_j's for the normal distribution was made through the use of the non-centrality parameter, θ, (Pearson and Hartley, 1951) where,

$$\theta = \sqrt{\frac{\Sigma n_j (\mu_j - \mu)^2}{J\sigma_e^2}}$$

Setting $\sigma_e^2 = 1$ and J = 4, and using probability of a Type I error equal to .05, the values of μ_j were found in a stepwise and a nonstepwise manner for both power values. Since equal sample size cases and unequal sample size cases would lead to different values of μ_j, the values were calculated for both equal and unequal sample sizes. Values of μ_j for sample sizes utilized in this study are presented in Table 1. The mean differences were used for the non-null situation for all variance conditions.

Four types of variances were established as follows: 1) equal variances (1:1:1:1); 2) unequal variances, stepwise (1:2:3:4); 3) unequal variances, with one small variance and three large variances, (SLLL, in the ratio of 1:4:4:4); 4) unequal variances with three small variances and one large variance (SSSL, in the ratio of 1:1:1:4). The average variance in all four cases was approximately equal to unity. (See Table 2)

The exponential distribution was derived by a method given by Lehman and Bailey (1968):

$$\text{for } f(t) = pe^{-pt}, \text{ with } p = 1,$$

$E(t) = 1/p=1$, and $var(t) = 1/p^2=1$. Pseudo-random exponential variables were generated by multiplying the negative of the mean, $-E(t) = -1$, times the natural logarithm of uniform random variates distributed on the unit interval (IBM, 1968). The exponential variates were then scaled so that the medians would be zero. The resulting skewed population had median equal to zero, a variance of σ_j^2, a skewness measure of $\gamma_1 = 2$, and a kurtosis measure of $\gamma_2 = 6$. For this distribution the medians have zero value under the null hypothesis, but the values of means are non-zero. For equal variances, the value of the means was .30685. For unequal variances, the mean for any group j is .30685 σ_j. In order to simulate null and non-null conditions in the exponential distribution the values of the medians for group j were identical to the values of μ_j used in the normal distribution as shown in Table 1. The variances were also identical to those used in the normal distribution as shown in Table 2.

Comparisons between the ANOVA F-test and the Kruskal-Wallis test for equal sample sizes and a total sample size of 28 were made on all four variance cases for both the normal distribution and the exponential

distribution. For three types of unequal sample sizes [(4,6,8,10), (4,8,8,8), (6,6,6,10)], total sample size of 28, comparisons were made on all four variance cases which were both positively and negatively related to sample sizes for both the normal distribution and the exponential distribution. In addition, for a total sample size of 68, two types of unequal sample sizes [(11,15,19,23), (11,19,19,19)] were used for comparisons of the F-test and the Kruskal-Wallis for only stepwise unequal variances, both positively and negatively related to sample sizes for both distributions. For each comparison of the F-test to the Kruskal-Wallis Test, 1,000 experiments were performed, where an experiment consisted of computation of each statistical test. The proportion of rejections in 1,000 experiments when there were no location differences was determined by comparing the empirical value of each test to five theoretical levels of significance, (.10, .05, .025, .01, .005) and each was referred to as empirical alpha. For each empirical alpha, four cases of empirical power were calculated by observing the proportion of rejections when each of the four types of differences in location was specified. A schematic which summarizes the conditions investigated can be found in the appendix. (See Table 13).

It should be noted that the equal sample size, equal variance case for the normal distribution was included in the present study for the purpose of establishing validity of the Monte Carlo method.

RESULTS

Equal Sample Sizes [7,7,7,7]

Normal population (Table 3): When all assumptions are met, both the ANOVA F-test and the Kruskal-Wallis test provide good approximations

to theoretical alpha. However, F-test approximations are somewhat closer to theoretical alpha than those of the Kruskal-Wallis test in the majority of cases. As is expected in this situation, the power of the F-test is higher than that of its nonparametric counterpart for all variance cases. In addition, the F-test displays only slight variability between power of stepwise mean differences and power of nonstepwise mean differences. The Kruskal-Wallis test, on the other hand, is greatly affected by the type of mean differences and usually displays a predominately higher power for stepwise differences than for nonstepwise differences. An illustration comparing power for the F-test and the Kruskal-Wallis test for a normal distribution with equal sample sizes for stepwise inequality of variances is shown in Figure 1, which is based on data from Table 3 in the Appendix.

Exponential Population (Table 4): When sampling from an exponential population with equal sample sizes, the empirical alphas of both tests for unequal variances cases more closely approximates theoretical alpha for the smaller values of alpha than when variances are equal. For all cases of variance inequality, empirical alphas of the ANOVA F-test and the Kruskal-Wallis test fell within two standard deviations of a proportion to theoretical alpha. Therefore, in terms of alpha alone, neither test could be recommended over the other.

With regard to power, discrimination between the two statistical procedures is evident. The Kruskal-Wallis test is generally a more powerful procedure than the ANOVA F-test for this case. When differences in means are stepwise, for every variance case and at all significance levels but one, the Kruskal-Wallis test is slightly more powerful than the F-test. Nonetheless, an interesting pattern of power values erupts

for nonstepwise mean differences. In the extreme tail regions (α=.01, α=.005), where Bradley (1968) indicates that parametric techniques would not be powerful, the ANOVA F-test is more powerful than the Kruskal-Wallis test in every instance of nonstepwise differences in means. This again illustrates the differential sensitivity (especially in remote tail regions) of the Kruskal-Wallis test to the type of location differences specified. It should be noted (see Figure 2 based on Table 4 in the Appendix) that as the level of significance gets smaller, the variability between the power of stepwise and nonstepwise mean differences becomes smaller for the ANOVA F-test and larger for the Kruskal-Wallis test.

Unequal Sample Sizes [4,6,8,10]

Normal population (Table 5): When variances are equal, there is little difference between the empirical alphas of the ANOVA F-test and those of the Kruskal-Wallis test. For all levels of significance, except α=.10, empirical values for equal variances are within sampling variability for both tests. With positively related sample sizes and variances, both the ANOVA F-test and the Kruskal-Wallis test are conservative (as expected from normal theory), with the ANOVA F-test exemplifying much closer approximations to all five theoretical alphas than does the Kruskal-Wallis test. There appears to be little difference in the empirical alpha values of either test as to the type of variance inequality for positive relationships with sample sizes. However, for negatively related sample sizes and variances, where normal theory expectations are liberal, the Kruskal-Wallis test exhibits conservativism for both stepwise unequal variances and for unequal variances in the ratio of 1:4:4:4 in the extreme tail regions. Despite the conservativism in some instances, the empirical alpha values for the Kruskal-Wallis test more closely approximate

theoretical alphas than do those of the F-test for all negatively related cases. In addition, the F-test makes more Type I errors than does the Kruskal-Wallis test for all negative relationships between sample sizes and variances. Solely on the basis of alpha approximantions, the Kruskal-Wallis test appears to be the better technique for negative relationships between sample sizes and variances, but a look at the empirical power values changes the perspective.

The power of both tests for nonstepwise mean differences is consistently less than the power for stepwise differences when sample sizes and variances are negatively related (See Figure 3, based on Table 5). Further, the power of the Kruskal-Wallis test is always lower than that of the ANOVA F-test for both types of mean differences, and reduces nearly to zero for nonstepwise differences established for small theoretical alphas (when α=.005; $(1-\beta)_{NSL}$ = .047, .026, .042). In the positively related sample sizes and variances cases where alpha approximations for both tests are good, the power of the Kruskal-Wallis test is still considerably less than the power of the ANOVA F-test. For a normal distribution with stepwise unequal sample sizes, if inequality of variances is suspected, the researcher should recognize the deficiency in power of the Kruskal-Wallis test.

Exponential Population (Table 6): Apparently, the addition of non-normality to unequal sample sizes of [4,6,8,10] has little effect on the empirical alphas of the ANOVA F-test and the Kruskal-Wallis test. These results from the exponential population are almost identical to the preceding section for a normal distribution for all variance cases. For stepwise inequality of variances positively related to sample sizes,

the empirical alphas of both tests are generally closer to theoretical alphas than those of the normal distribution, and for negative relationships they are slightly farther away from theoretical alphas than those of the normal distribution. The same unexpected conservativism occurs for the Kruskal-Wallis test when sample sizes and variances are negatively related for only two cases of theoretical alpha.

Sampling from an exponential population does have a different effect on power. For positively related sample sizes and variances, the power of the Kruskal-Wallis test is comparable to or larger than the power of the ANOVA F-test in all cases of unequal variances for both types of mean differences. Neither the Kruskal-Wallis test nor the ANOVA F-test shows any particular trend in power between stepwise and nonstepwise mean differences for the positively related cases. But with negatively related cases of sample sizes and variances, the power of the Kruskal-Wallis test once again behaves strangely (See Figure 4 based on Table 6). When nonstepwise mean differences are specified, they are, as shown in previous results, a deterrent to the use of the Kruskal-Wallis test in negatively related cases for small theoretical alphas. As seen in Table 6, the power of the Kruskal-Wallis test approaches zero for nonstepwise differences in all variance cases. Throughout the entire range of theoretical alphas, stepwise mean differences generate larger power values for both tests than do nonstepwise mean differences in the negatively related sample sizes and variances cases. For small theoretical alphas, the Kruskal-Wallis test, even for stepwise mean differences, produces power values less than those of the ANOVA F-test in a few instances for the negatively related cases. (See Table 6; $\neq$ n's, $\neq \sigma_e^2$; SSSL; α = .01, .005) The power of the ANOVA F-test for both stepwise and nonstepwise differences does remain fairly stable for all levels of theoretical alpha. Due to the instability

and lack of power for the Kruskal-Wallis test with negatively related sample sizes and variances, the researcher should exercise extreme caution in utilizing this nonparametric procedure.

Unequal Sample Sizes [4,8,8,8] [6,6,6,10]

Due to the multiplicity of data, and the similarity of results, these two cases of unequal sample sizes will be grouped together in a generalized commentary.

Normal Population (Tables 7 and 9): When variances are equal, approximations to theoretical alpha for both tests and both sample sizes are good. For the positively related cases of sample sizes and variance inequality, both tests are conservative in most cases. The ANOVA F-test and the Kruskal-Wallis test do show an unexpected distinct reversal from normal theory in a few cases [(Table 7; $\neq$ n's, $\neq \sigma_e^2$; SSSL) (Table 9; $\neq$ n's, $\neq \sigma_e^2$; SLLL)], but all are within sampling variability of theoretical alpha. For the negatively related cases, the Kruskal-Wallis test is usually conservative when a liberal test is expected [(Table 7; $\neq$ n's, $\neq \sigma_e^2$; Stepwise, SSSL, SLLL) (Table 9; $\neq$ n's, $\neq \sigma_e^2$; Stepwise, SSSL, SLLL)]. There are scattered instances where the ANOVA F-test is also unexpectedly conservative, but in all cases of this type the ANOVA F-test is less conservative than is the Kruskal-Wallis test. These reversals from normal theory could very well be a function of the type of sample size inequality in conjunction with the type of variance inequality. Unanticipated reversals of this type create a problem for the researcher and further investigation in this area is needed. At best, if the Kruskal-Wallis test is used, the researcher with sample sizes of this type should expect a conservative test no matter what type of variance inequality is suspected. In using the ANOVA F-test, normal theory holds in the majority of cases.

When variances are equal and when variances are positively related to sample sizes, there is almost no difference between power of stepwise mean differences and nonstepwise mean differences for either test. Generally, for these variance cases, power of the ANOVA F-test is comparable to or better than that of the Kruskal-Wallis test. Certainly there is no indication that nonparametric power is significantly more enticing than parametric power or vice versa.

In the negatively related sample sizes and variances cases, there is, as shown throughout the results, an obvious difference in power between the two statistical procedures. Stepwise mean differences produce higher power than nonstepwise mean differences for both tests, with the ANOVA F-test having better power in practically all cases. Empirical power values of the Kruskal-Wallis test for nonstepwise mean differences fall to almost zero for small set alphas, and large nonstepwise differences reduce to almost the same empirical power values as those produced by small nonstepwise mean differences. In other words, whether large or small nonstepwise mean differences are specified, the power is comparable. Apparently, large nonstepwise mean differences for the Kruskal-Wallis test are severely affected in terms of power for small alphas when variances and sample sizes are negatively related (See Figure 5 based on Table 7). Overall, the ANOVA F-test seems to be a better technique in terms of power.

Exponential Population (Tables 8 and 10): When variances are equal, the Kruskal-Wallis test is slightly conservative for all theoretical alpha values, with one exception, for both sample sizes. The ANOVA F-test, however, provides empirical alphas that are very close to theoretical alphas for sample sizes of [4,8,8,8] and are slightly conservative for

sample sizes of [6,6,6,10]. For small values of theoretical alpha, the ANOVA F-test again shows a reversal from normal theory when variances and sample sizes are positively related. [(Table 8; ≠ n's, ≠ σ_e^2; Stepwise, SLLL) (Table 10; ≠ n's, ≠ σ_e^2; SSSL)] The Kruskal-Wallis test is conservative in all cases of positively related sample sizes and variances. Empirical alphas for the negatively related cases for both of these sample sizes are considerably closer to theoretical alpha than were those of the stepwise unequal sample size cases ([4,6,8,10]; Table 6). In fact, empirical alphas of both tests are consistently within two standard deviations of a proportion difference of theoretical alpha.

As far as power is concerned, the Kruskal-Wallis test displays the same lack of power for nonstepwise mean differences for small values of theoretical alpha that it has displayed in previous cases. (See Figure 6 based on Table 8) For stepwise mean differences, there is only a small difference between the power of the ANOVA F-test and the power of the Kruskal-Wallis test, with the power of the Kruskal-Wallis test being slightly higher in most cases. Nevertheless, the undependability of empirical power of the Kruskal-Wallis test in certain situations prevents a recommendation for use as a more powerful technique than the ANOVA F-test.

Unequal Sample Sizes [11,15,19,23] [11,19,19,19]

Normal Population (Table 11): For positively related sample sizes and stepwise unequal variances, empirical alphas of the ANOVA F-test and the Kruskal-Wallis test are generally closer to set alphas when sample sizes are [11,19,19,19] than for the stepwise unequal sample sizes [11,15,19,23]. When sample sizes and variances are negatively related, the stepwise increment in sample sizes provides slightly closer approximations to set alphas, but the differences in alphas for the two sample sizes are negligible.

As is usually the case, the two statistical tests are conservative in all cases for positive relationships between sample sizes and variances. However, for negative relationships in a few cases (α=.01), the Kruskal-Wallis test is found to be slightly conservative when a liberal test is expected, but the empirical alphas are all within sampling variability. The ANOVA F-test for negative relationships with sample sizes and variances, tends to make a few too many Type I errors in all but one case. Upon comparing empirical alphas of the two statistical procedures, one does not significantly outperform the other for the cases investigated using these sample sizes.

In terms of power, stepwise mean differences produce higher empirical power than nonstepwise mean differences for both statistical tests using both types of sample size inequality. Apparently using a larger total sample size has an effect on the Kruskal-Wallis test since the power differential between the Kruskal-Wallis test and the ANOVA F-test for negatively related cases, nonstepwise mean differences for small set alphas, is much less than was seen in the previous results. The empirical power of the Kruskal-Wallis test is still less than the empirical power of the F-test in these remote tail regions, but there is not as great a difference. For the majority of cases, the ANOVA F-test is the more powerful technique but the empirical power of the Kruskal-Wallis test is comparable for large values of set alpha [.10, .05].

Exponential Population (Table 12): Similar to findings with the normal distribution for positively related sample sizes and variances, empirical alphas of both tests for sample sizes of [11,19,19,19] provide closer approximations to theoretical alpha than for sample sizes of [11,15,19,23],

except when theoretical α=.005. Empirical alphas of both the Kruskal-Wallis test and the F-test are generally less than theoretical alpha, with the exception of one case, when α=.005 for the Kruskal-Wallis test. For negatively related sample sizes and variances, when sample sizes are unequal in a stepwise manner, both tests are liberal as expected. However, normal theory liberalism of alpha does not hold for unequal sample sizes of [11,19,19,19]. Every empirical alpha of the ANOVA F-test in this case is less than theoretical alpha, thus illustrating a gross deviation from normal expectancy. The Kruskal-Wallis test is also affected, but only shows reversals for three out of the five theoretical alpha values. The unusual behavior of the ANOVA F-test for the larger sample size (N=68 as opposed to N=28) indicates that perhaps for large unequal sample sizes when non-normality is suspected along with unequal variances, the Kruskal-Wallis test might be the preferred procedure in terms of approximating theoretical alpha.

Empirical power for both tests is generally higher for stepwise mean differences than for nonstepwise mean differences. There is less differences between empirical power based on stepwise differences and that based on nonstepwise differences for these larger sample sizes than for the smaller sample sizes previously discussed. For unequal sample sizes of [11,19,19,19], the Kruskal Wallis test with negatively related sample sizes and variances, even for nonstepwise mean differences, demonstrates higher power than that of the ANOVA F-test. Overall, the Kruskal-Wallis test is generally a more powerful technique than the F-test for these sample sizes.

SUMMARY AND CONCLUSIONS

The Kruskal-Wallis test has often been recommended for use in lieu of the ANOVA F-test when assumptions of the parametric procedure are

violated. On the other hand, research has shown the ANOVA F-test to be relatively robust to a variety of assumption violations. To better guide the researcher, the present study empirically compared the robustness of the Kruskal-Wallis test and the ANOVA F-test in terms of probability of a Type I error and power under an assortment of varying conditions.

Upon comparing empirical alphas of the two statistical procedures, it was found for equal variances and for positively related sample sizes and variances that the ANOVA F-test provided closer approximations to theoretical alphas in the majority of cases for both normal and exponential distributions for all combinations of sample size inequality when N=28. When sample sizes were equal, the two procedures were comparable with the Kruskal-Wallis test providing slightly better alpha approximations for the exponential distribution and the ANOVA F-test providing slightly better alpha approximations for the normal distribution. When sample sizes and variances were negatively related, the empirical alphas of the Kruskal-Wallis test generally more closely approximated theoretical alphas for both distributions and all combinations of sample size inequality but one ([10,6,6,6], Normal Population). For larger unequal sample sizes (N=68), the Kruskal-Wallis test generally demonstrated slightly closer approximations to theoretical alpha for the normal distribution and the exponential distribution.

For both statistical tests when sample sizes and variances were unequal, reversals from normal theory alphas were noted in some cases. Both tests were conservative for negative relationships and liberal for positive relationships in a few cases for the normal distribution with sample sizes of [4,8,8,8] and [6,6,6,10]. When a liberal test

was expected, the Kruskal-Wallis test was found to be conservative in the extreme tail regions for sample sizes of [4,6,8,10], [11,15,19,23] and [11,19,19,19]. Also, the F-test was found to be unexpectedly conservative in a few cases for sample sizes of [4,8,8,8], [6,6,6,10] and in all cases for sample sizes of [11,19,19,19] from the normal distribution, when sample sizes were negatively related to variances. These reversals were not found to be particularly illustrative of a trend toward poor estimates of theoretical alpha but rather simply indicative of specific instances of deviation from normal theory expectancy and warrant mention.

On the basis of alpha approximations alone, the F-test provided the best approximations to theoretical alpha in the majority of cases investigated in this study with the Kruskal-Wallis test demonstrating comparable or better approximations in specific cases.

Power comparisons indicated that when sampling from a normal distribution with both positively and negatively related sample sizes and variances, the empirical power of the ANOVA F-test was higher than that of the Kruskal-Wallis test. When positive relationships of sample sizes and variances were present for the exponential distribution the Kruskal-Wallis test generally demonstrated slightly higher power than that of the F-test. However, for negative relationships, with sampling from an exponential distribution the Kruskal-Wallis test consistently displayed higher power than the F-test only when stepwise mean differences were specified, and even then there was little difference between empirical power of the two procedures. For nonstepwise mean differences, the Kruskal-Wallis test was found to be extremely sensitive. In fact for small theoretical alphas, the power of the Kruskal-Wallis test

diminished almost to zero for negative relationships of sample sizes and variances for both distributions. An interesting point concerning this lack of power for the Kruskal-Wallis test for nonstepwise mean differences was it's consistency. The empirical power values were found to be consistently low across distributions, across variance patterns, and across sample size patterns for N=28. Due to the lack of power of the Kruskal-Wallis test in these cases, a recommendation for use at small alphas cannot be given on the basis of this research. Further, most researchers have no idea about the type of mean differences present in their data, and to expect them to be stepwise in nature is unreasonable. Therefore, a researcher utilizing small unequal sample sizes suspecting inequality of variances could be almost totally without power if testing hypotheses at α=.01 or smaller using the Kruskal-Wallis test. The F-test was found to generally have less power with nonstepwise mean differences than for stepwise mean differences in the negatively related cases but power for the F-test was much higher than that of the Kruskal-Wallis test and showed no indication of diminishing to zero. It should be noted, however, that for the larger unequal sample sizes investigated (N=68), the Kruskal-Wallis test was competitive in terms of power for nonstepwise mean differences if the distribution was exponential. But the lack of power was still evident for the Kruskal-Wallis test even for larger sample sizes for normal distributions. Since only a few cases were investigated for N=68, it would be difficult to generalize concerning the stability or instability of empirical power values for either statistical procedure.

In conclusion, when normality and/or homogeneity of variance is doubtful, the ANOVA F-test was found to be the recommended procedure

for testing hypotheses, especially at significance levels of .01 or less. This recommendation is based primarily on the instability of power for the Kruskal-Wallis test. If a researcher is unconcerned with power, the Kruskal-Wallis test could be recommended as being competitive to the F-test in terms of empirical alphas.

REFERENCES

BONEAU, C. A. (1960). The effects of violations of assumptions underlying the t-test. Psy. Bull. 57, 49-64.

BOX, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Ann. math. Statist. 26, 290-302.

BOX, G. E. P. & ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departure from assumptions. J.R. Statist. Soc. B. 17, 1-34.

BOX, G. E. P. & MULLER, M. E. (1958). A note on the generation of random normal deviates. Ann. math. Statist. 29, 610-611.

BRADLEY, J. V. (1963). Studies in Research Methodology IV: A sampling Study of the Central Limit Theorem and the Robustness of One-Sample Parametric Tests. AMRL Technical Document Report, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

BRADLEY, J. V. (1966). Studies in Research Methodology VII: The Central Limit Effect for Two Dozen Populations and its Correlation With Population Moments. Behavioral Science Laboratory, Wright-Patterson Air Force Base, Ohio.

BRADLEY, J. V. (1968). Distribution-Free Statistical Tests. London: Prentice-Hall.

COCHRAN, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics 3, 22-38.

DAVID, F. N. & JOHNSON, N. L. (1951). The effect of non-normality on the power function of the F-test in the analysis of variance. Biometrika 38, 43-57.

DONALDSON, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. J.A.S.A. 63, 660-676.

FREIR, B. & TOOTHAKER, L. E. (1972) An empirical comparison of the ANOVA F-test, the Kruskal-Wallis test, and the normal scores test under violation of assumptions. Unpublished manuscript.

GAMES, P. A. & LUCAS, P. A. (1966). Power of the analysis of variance of independent groups on nonnormal and normally transformed data. Ed. Psy. Meas. 26, 311-327.

GLASS, G. V. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. R.E.R. 42, 237-288.

HODGES, J. L. & LEHMAN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. Proc. Fourth Berkely Symp. Math. Statist. & Prob. Berkely and Los Angeles: University of California Press.

HSU, P. L. (1938). Contribution to the theory of "student's" t-test as applied to the problem of two samples. Statist. Res. Mem. 2, 1-24.

INTERNATIONAL BUSINESS MACHINES CORPORATION (1968). System/360, Scientific Subroutine Package. Programmer's Manual H20-0205-3, 77.

KENDALL, M. G. & STUART, A. (1961). The Advanced Theory of Statistics. London: Griffin and Company.

KLOTZ, J. (1963). Small sample power and efficiency for the one sample Wilcoxon and normal scores test. Ann. math. Statist. 34, 624-632.

KRUSKAL, W. H. (1952). A nonparametric test for the several sample problem. Ann. math. Statist. 23, 525-540.

LEHMAN, R. S. & BAILEY, D. E. (1968). Digital Computing. New York; John Wiley.

LINDQUIST, E. F. (1953). Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Co.

LUNNEY, G. H. (1968). A Monte Carlo investigation of basic analysis of variance models when the dependent variable is a Bernoulli variable. Dissertation Abstracts, Ann Arbor, Michigan (order no. 68-17, 696).

LUNNEY, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. J. Ed. Meas. 7, 263-269.

McSWEENEY, M. (1967). An empirical study of two proposed nonparametric tests for main effects and interaction. (Doctor's thesis, University of California, Berkeley.)

McSWEENEY, M. & PENFIELD, D. (1969). The normal scores test for the c-sample problem. Br. J. math. statist. Psychol. 22, 177-192.

NORTON, D. W. (1952). An empirical investigation of some effects of non-normality and heterogeneity of the F-distribution. (Doctor's thesis, State University of Iowa).

PEARSON, E. S. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. Biometrika 21, 259-286.

PEARSON, E. S. (1931). The analysis of variance in cases of non-normal variation. Biometrika 23, 114-133.

PEARSON, E. S. & HARTLEY, H. O. (1958). Chart of the power function for analysis of variance tests, derived from the non-central F-distribution. Biometrika 38. 112-130.

PENFIELD, D. & McSWEENEY, M. E. (1968). The normal scores test for the two-two-sample problem. Psy. Bull. 69, 183-191.

PRATT, J. W. (1964). Robustness of some procedures for the two-sample location problem. J.A.S.A. 59, 665-680.

PRUI, M. L. (1964). Asymptotic efficiency of a class of c-sample tests. Ann. math. Statist. 35, 102-121.

RIDER, P. R. (1929). On the distribution of the ratio of mean to standard deviation in small samples from non-normal populations. Biometrika 21, 124-143.

SCHEFFE, H. (1959). The Analysis of Variance. New York: John Wiley.

SRIVASTAVA, A.B.L. (1958). Effect of non-normality on the power function of t-test. Biometrika 45, 421-30.

TERRY, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. Ann. math. Statist. 23, 346-366.

TOOTHAKER, L. E. (1972). An empirical investigation of the permutation t-test. Br. J. math. statis. Psychol. 25, 83-94.

TOOTHAKER, L. E. (1973). An empirical investigation of the effect of unequal variances on the permutation t-test. Unpublished manuscript.

VAN DER LAAN, P. (1964). Exact power of some rank tests. Publ. Inst. Statist. Univ. Paris 13, 211-234.

APPENDIX A

TABLE 1

Values of n_j and μ_j

j	n_j	$\mu_{j_{SL}}$	$\mu_{j_{NSL}}$	$\mu_{j_{SS}}$	$\mu_{j_{NSS}}$
1	7	0	.577	0	.419
2	7	.67	.577	.487	.419
3	7	1.34	.577	.974	.419
4	7	2.01	2.308	1.461	1.676
1	4	0	.521	0	.379
2	6	.707	.521	.513	.379
3	8	1.414	.521	1.026	.379
4	10	2.121	2.084	1.539	1.516
1	4	0	.553	0	.402
2	8	.727	.553	.529	.402
3	8	1.454	.553	1.058	.402
4	8	2.181	2.212	1.587	1.608
1	6	0	.521	0	.379
2	6	.646	.521	.469	.379
3	6	1.292	.521	.938	.379
4	10	1.938	2.084	1.407	1.516
1	11	0	.315	0	.236
2	15	.415	.315	.310	.236
3	19	.830	.315	.620	.236
4	23	1.245	1.260	.930	.944
1	11	0	.332	0	.249
2	19	.426	.332	.319	.249
3	19	.852	.332	.638	.249
4	19	1.278	1.328	.957	.996

n_j = number of observations per treatment level

$\mu_{j_{SL}}$ = a stepwise difference such that power was approximately .86 for α=.05 for the normal distribution

$\mu_{j_{NSL}}$ = a non-stepwise difference (1:1:1:4) such that power was approximately .86 for α=.05 for the normal distribution

$\mu_{j_{SS}}$ = a stepwise difference such that power was approximately .60 for α=.05 for the normal distribution

$\mu_{j_{NSS}}$ = a non-stepwise difference (1:1:1:4) such that power was approximately .60 for α=.05 for the normal distribution

TABLE 2

Values of Variances

j	$=\sigma^2$	Stepwise	SSSL*	SLLL**
1	1.0	.4	.5714	.30769
2	1.0	.8	.5714	1.23076
3	1.0	1.2	.5714	1.23076
4	1.0	1.6	2.2857	1.23076

* SSSL - These variances are approximately in the ratio 1:1:1:4

** SLLL - These variances are approximately in the ratio 1:4:4:4

TABLE 3

Normal Population-Equal Sample Sizes (7,7,7,7)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large;NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$= \sigma_e^2$	α	.096	.043	.019	.007	.002	.090	.035	.014	.004	.001
	$1-\beta_{SL}$	.940	.872	.765	.640	.526	.921	.817	.702	.505	.361
	$1-\beta_{NSL}$	.936	.875	.803	.662	.569	.905	.832	.703	.480	.313
	$1-\beta_{SS}$	.719	.601	.485	.325	.234	.676	.538	.404	.226	.121
	$1-\beta_{NSS}$	.749	.613	.477	.331	.233	.697	.546	.376	.196	.104
$\neq \sigma_e^2$ (stepwise)	α	.104	.063	.042	.018	.009	.111	.054	.024	.005	.004
	$1-\beta_{SL}$	.957	.900	.814	.675	.579	.951	.880	.754	.582	.421
	$1-\beta_{NSL}$	.899	.829	.756	.645	.552	.831	.713	.594	.425	.275
	$1-\beta_{SS}$	.719	.590	.459	.322	.241	.704	.547	.401	.240	.150
	$1-\beta_{NSS}$	.678	.551	.429	.301	.231	.604	.430	.302	.172	.102
$\neq \sigma_e^2$(SSSL)	α	.103	.056	.028	.014	.009	.091	.035	.014	.007	.003
	$1-\beta_{SL}$	.921	.855	.774	.668	.557	.923	.836	.726	.536	.419
	$1-\beta_{NSL}$	.856	.798	.719	.601	.526	.753	.642	.515	.339	.236
	$1-\beta_{SS}$	.693	.576	.449	.318	.258	.692	.540	.384	.232	.145
	$1-\beta_{NSS}$	.645	.555	.459	.350	.276	.536	.397	.260	.147	.084
$\neq \sigma_e^2$(SLLL)	α	.098	.060	.025	.007	.004	.101	.051	.017	.007	.003
	$1-\beta_{SL}$	.962	.910	.823	.684	.578	.958	.907	.794	.593	.437
	$1-\beta_{NSL}$	.931	.860	.781	.678	.570	.898	.797	.682	.496	.334
	$1-\beta_{SS}$	.735	.588	.446	.292	.206	.758	.586	.406	.222	.122
	$1-\beta_{NSS}$	.709	.581	.455	.317	.247	.666	.504	.361	.191	.104

TABLE 4

Exponential Population-Equal Sample Sizes (7,7,7,7)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-Test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$= \sigma_e^2$	α	.094	.045	.017	.005	.000	.099	.052	.014	.002	.001
	$1-\beta_{SL}$	.919	.867	.793	.685	.616	.959	.910	.860	.715	.599
	$1-\beta_{NSL}$	.936	.895	.817	.722	.630	.980	.938	.852	.638	.427
	$1-\beta_{SS}$	.768	.643	.532	.397	.329	.855	.749	.615	.445	.315
	$1-\beta_{NSS}$	.744	.640	.531	.388	.290	.874	.739	.573	.341	.180
$\neq \sigma_e^2$ (stepwise)	α	.094	.054	.025	.009	.006	.109	.052	.024	.008	.004
	$1-\beta_{SL}$	.988	.956	.902	.821	.742	.995	.984	.963	.898	.826
	$1-\beta_{NSL}$	.961	.931	.878	.794	.712	.991	.967	.886	.696	.501
	$1-\beta_{SS}$	.878	.768	.638	.453	.350	.950	.879	.790	.607	.474
	$1-\beta_{NSS}$	.804	.698	.583	.446	.355	.890	.777	.615	.370	.200
$\neq \sigma_e^2$ SSSL	α	.116	.059	.026	.010	.009	.104	.054	.024	.006	.004
	$1-\beta_{SL}$	.974	.951	.906	.838	.761	.988	.974	.946	.872	.800
	$1-\beta_{NSL}$	.986	.965	.924	.840	.764	.997	.977	.915	.763	.544
	$1-\beta_{SS}$	.849	.748	.637	.488	.392	.915	.834	.740	.588	.453
	$1-\beta_{NSS}$	.860	.763	.658	.482	.375	.899	.797	.636	.359	.215
$\neq \sigma_e^2$ SLLL	α	.092	.047	.023	.010	.008	.103	.052	.022	.009	.004
	$1-\beta_{SL}$	.976	.959	.916	.832	.752	.997	.988	.966	.919	.846
	$1-\beta_{NSL}$	.946	.894	.822	.725	.640	.985	.951	.850	.645	.439
	$1-\beta_{SS}$	.884	.794	.671	.517	.389	.967	.918	.834	.656	.512
	$1-\beta_{NSS}$	.770	.671	.558	.429	.338	.903	.781	.612	.372	.203

TABLE 5

Normal Population-Unequal Sample Sizes, (4, 6, 8, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.127	.052	.030	.012	.007	.126	.050	.022	.005	.001
	$1-\beta_{SL}$	.927	.860	.791	.652	.545	.906	.821	.685	.488	.341
$= \sigma_e^2$	$1-\beta_{NSL}$	.922	.859	.781	.651	.559	.909	.833	.718	.545	.399
	$1-\beta_{SS}$	.724	.580	.464	.318	.239	.680	.523	.382	.199	.100
	$1-\beta_{NSS}$	.732	.600	.467	.314	.231	.699	.538	.384	.220	.118
$\neq$ n's	α	.078	.040	.017	.005	.003	.081	.032	.011	.003	.002
	$1-\beta_{SL}$	.931	.865	.780	.644	.512	.933	.840	.706	.490	.314
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.901	.828	.738	.597	.501	.876	.786	.664	.477	.339
	$1-\beta_{SS}$	.691	.529	.401	.260	.183	.670	.490	.338	.179	.089
positively related	$1-\beta_{NSS}$	.655	.530	.404	.276	.207	.622	.468	.329	.199	.118
$\neq$ n's	α	.163	.082	.048	.023	.010	.112	.057	.028	.004	.001
	$1-\beta_{SL}$	.944	.906	.849	.752	.672	.920	.860	.759	.597	.443
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.745	.636	.549	.441	.363	.595	.460	.308	.112	.047
	$1-\beta_{SS}$	.802	.675	.575	.429	.337	.747	.593	.450	.262	.154
negatively related	$1-\beta_{NSS}$	.541	.430	.335	.228	.168	.412	.277	.156	.050	.023
$\neq$ n's	α	.077	.041	.017	.008	.005	.078	.036	.007	.002	.000
	$1-\beta_{SL}$	.906	.841	.759	.618	.517	.897	.807	.688	.497	.353
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.906	.829	.740	.613	.527	.872	.778	.653	.476	.337
	$1-\beta_{SS}$	.699	.552	.426	.284	.208	.665	.510	.348	.185	.105
positively related	$1-\beta_{NSS}$	.676	.556	.440	.291	.215	.632	.492	.343	.187	.113
$\neq$ n's	α	.188	.120	.085	.050	.038	.126	.076	.034	.007	.002
	$1-\beta_{SL}$	.967	.928	.874	.782	.706	.953	.884	.806	.662	.512
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.702	.616	.527	.455	.379	.539	.403	.271	.085	.026
	$1-\beta_{SS}$	.798	.681	.571	.469	.377	.752	.607	.459	.291	.193
negatively related	$1-\beta_{NSS}$	.509	.421	.338	.264	.201	.356	.261	.162	.055	.016
$\neq$ n's	α	.099	.045	.025	.008	.007	.088	.038	.015	.005	.003
	$1-\beta_{SL}$	.937	.868	.781	.639	.530	.919	.846	.700	.498	.325
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.920	.863	.783	.658	.550	.904	.820	.716	.532	.397
	$1-\beta_{SS}$	.679	.531	.412	.255	.166	.668	.503	.342	.173	.101
positively related	$1-\beta_{NSS}$	.680	.550	.421	.280	.189	.660	.497	.350	.180	.100
$\neq$ n's	α	.148	.079	.058	.024	.015	.122	.059	.026	.010	.002
	$1-\beta_{SL}$	.966	.924	.875	.772	.698	.953	.905	.819	.625	.481
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.756	.638	.539	.413	.318	.677	.502	.315	.109	.042
	$1-\beta_{SS}$	.767	.652	.547	.395	.304	.752	.601	.443	.255	.158
negatively related	$1-\beta_{NSS}$	.507	.374	.272	.167	.122	.423	.270	.142	.053	.024

TABLE 6

Exponential Population-Unequal Sample Sizes, (4, 6, 8, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large;NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
	α	.094	.048	.029	.017	.010	.101	.043	.016	.010	.004
$\neq$ n's	$1-\beta_{SL}$	.941	.876	.811	.708	.627	.957	.910	.840	.732	.617
	$1-\beta_{NSL}$	.941	.899	.824	.701	.605	.974	.949	.888	.737	.581
$= \sigma_e^2$	$1-\beta_{SS}$	.745	.626	.538	.384	.295	.846	.738	.634	.458	.332
	$1-\beta_{NSS}$	.768	.640	.544	.394	.309	.886	.796	.646	.465	.302
$\neq$ n's	α	.075	.029	.020	.010	.006	.081	.034	.015	.006	.002
	$1-\beta_{SL}$	.946	.902	.847	.739	.644	.981	.951	.912	.808	.675
$\neq \sigma^2$ (stepwise)	$1-\beta_{NSL}$	.921	.871	.798	.682	.609	.977	.934	.867	.739	.579
	$1-\beta_{SS}$	.799	.662	.547	.388	.294	.886	.800	.694	.537	.370
positively related	$1-\beta_{NSS}$	.780	.663	.541	.394	.300	.887	.798	.665	.463	.299
$\neq$ n's	α	.165	.105	.054	.032	.019	.156	.084	.043	.012	.004
	$1-\beta_{SL}$	.991	.976	.949	.902	.847	.997	.992	.975	.925	.858
$\neq \sigma^2$ (stepwise)	$1-\beta_{NSL}$	.854	.763	.675	.562	.464	.866	.650	.382	.131	.055
	$1-\beta_{SS}$	.911	.832	.739	.611	.520	.959	.894	.828	.639	.480
negatively related	$1-\beta_{NSS}$	.631	.517	.418	.295	.222	.619	.407	.229	.087	.029
= n's	α	.064	.036	.019	.010	.003	.075	.035	.016	.001	.000
	$1-\beta_{SL}$	.964	.921	.867	.777	.688	.979	.959	.917	.831	.722
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.968	.931	.872	.775	.687	.990	.971	.917	.816	.680
	$1-\beta_{SS}$	.796	.670	.561	.404	.306	.878	.788	.675	.505	.385
positively related	$1-\beta_{NSS}$	.808	.714	.574	.402	.305	.914	.823	.707	.494	.344
$\neq$ n's	α	.179	.103	.059	.031	.020	.130	.070	.043	.012	.006
	$1-\beta_{SL}$	.976	.956	.924	.872	.809	.991	.977	.954	.886	.788
$\neq \sigma^2$ (SSSL)	$1-\beta_{NSL}$	.871	.803	.713	.583	.507	.873	.660	.393	.131	.045
	$1-\beta_{SS}$	.876	.814	.736	.607	.506	.921	.857	.761	.588	.446
negatively related	$1-\beta_{NSS}$	.670	.549	.439	.318	.235	.583	.341	.159	.054	.018
$\neq$ n's	α	.071	.030	.016	.010	.005	.083	.033	.017	.009	.004
	$1-\beta_{SL}$	.960	.916	.848	.729	.638	.976	.954	.909	.801	.678
$\neq \sigma^2$ (SLLL)	$1-\beta_{NSL}$	.944	.889	.818	.716	.600	.983	.945	.886	.732	.586
	$1-\beta_{SS}$	.798	.682	.548	.409	.304	.902	.809	.712	.526	.400
positively related	$1-\beta_{NSS}$	.745	.628	.487	.350	.282	.885	.774	.626	.408	.264
$\neq$ n's	α	.140	.085	.046	.021	.015	.142	.082	.034	.011	.002
	$1-\beta_{SL}$	.987	.965	.929	.869	.799	1.000	.995	.979	.913	.843
$\neq \sigma^2$ (SLLL)	$1-\beta_{NSL}$	.825	.731	.630	.510	.408	.875	.663	.399	.157	.077
	$1-\beta_{SS}$	.913	.843	.727	.595	.466	.970	.919	.841	.645	.463
negatively related	$1-\beta_{NSS}$	.622	.502	.388	.273	.210	.659	.412	.232	.086	.038

TABLE 7

Normal Population-Unequal Sample Sizes, (4, 8, 8, 8)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large;NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases (= σ_e^2; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.092	.042	.021	.008	.005	.090	.037	.016	.006	.002
$\neq \sigma_e^2$	$1-\beta_{SL}$	.935	.865	.788	.664	.566	.907	.836	.702	.518	.353
	$1-\beta_{NSL}$	.937	.878	.798	.675	.562	.921	.838	.705	.516	.359
	$1-\beta_{SS}$	.745	.620	.484	.342	.244	.703	.557	.398	.209	.117
	$1-\beta_{NSS}$	.706	.574	.465	.312	.228	.677	.530	.364	.198	.111
$\neq$ n's	α	.086	.042	.025	.006	.004	.082	.034	.014	.007	.003
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.945	.878	.798	.658	.548	.919	.842	.725	.511	.351
	$1-\beta_{NSL}$	.946	.874	.794	.669	.553	.915	.820	.687	.503	.346
	$1-\beta_{SS}$	.709	.573	.456	.300	.220	.681	.516	.355	.193	.109
positively related	$1-\beta_{NSS}$	.708	.575	.465	.327	.234	.654	.519	.363	.207	.125
$\neq$ n's	α	.124	.061	.031	.013	.011	.105	.047	.018	.007	.002
$\neq \sigma^2$ (stepwise)	$1-\beta_{SL}$	.956	.903	.835	.708	.616	.945	.855	.760	.567	.409
	$1-\beta_{NSL}$	.769	.670	.553	.418	.303	.655	.489	.313	.090	.028
	$1-\beta_{SS}$	.781	.667	.542	.384	.298	.747	.594	.437	.270	.167
negatively related	$1-\beta_{NSS}$	.554	.420	.319	.208	.142	.461	.295	.157	.045	.014
$\neq$ n's	α	.106	.054	.027	.011	.004	.111	.046	.020	.003	.001
$\neq \sigma^2$ (SSSL)	$1-\beta_{SL}$	.949	.889	.817	.706	.609	.932	.848	.747	.545	.401
	$1-\beta_{NSL}$	.916	.853	.778	.666	.568	.881	.789	.662	.483	.328
	$1-\beta_{SS}$	.713	.577	.463	.329	.234	.671	.514	.382	.210	.129
positively related	$1-\beta_{NSS}$	.709	.576	.455	.309	.230	.640	.495	.345	.189	.106
$\neq$ n's	α	.126	.070	.043	.024	.013	.094	.050	.021	.003	.000
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{SL}$	.951	.911	.852	.742	.659	.935	.859	.764	.592	.440
	$1-\beta_{NSL}$	.784	.699	.592	.471	.388	.636	.501	.321	.097	.030
	$1-\beta_{SS}$	.782	.677	.553	.414	.314	.732	.589	.444	.257	.151
negatively related	$1-\beta_{NSS}$	.563	.448	.351	.251	.179	.455	.290	.163	.050	.018
$\neq$ n's	α	.080	.042	.021	.010	.004	.080	.033	.011	.005	.000
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{SL}$	.938	.891	.806	.685	.571	.933	.849	.722	.541	.371
	$1-\beta_{NSL}$	.935	.870	.783	.663	.573	.917	.809	.699	.524	.358
	$1-\beta_{SS}$	.711	.563	.427	.261	.196	.684	.526	.346	.178	.092
positively related	$1-\beta_{NSS}$	.691	.555	.425	.295	.218	.664	.497	.341	.184	.098
$\neq$ n's	α	.092	.049	.029	.007	.004	.092	.042	.018	.006	.002
$\neq \sigma^2$ (SLLL)	$1-\beta_{SL}$	.966	.916	.837	.711	.600	.960	.897	.768	.568	.414
	$1-\beta_{NSL}$	.788	.656	.543	.408	.315	.702	.512	.311	.101	.039
	$1-\beta_{SS}$	.784	.671	.560	.383	.260	.766	.629	.469	.236	.125
negatively related	$1-\beta_{NSS}$	.556	.408	.289	.172	.120	.474	.291	.158	.044	.015

TABLE 8

Exponential Population-Unequal Sample Sizes, (4, 8, 8, 8)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power (1-β). Power is based on four cases of mean differences (SL-stepwise, large; NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases (= σ_e^2; ≠ σ_e^2, stepwise; ≠ σ_e^2, SSSL, approximately in the ratio of 1:1:1:4; and ≠ σ_e^2, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.01	.05	.025	.01	.005	.01	.05	.025	.01	.005
≠ n's	α	.095	.051	.026	.012	.008	.095	.042	.020	.004	.000
2	$1-\beta_{SL}$	.922	.868	.809	.712	.607	.961	.922	.854	.729	.584
= σ_e	$1-\beta_{NSL}$	.950	.897	.842	.722	.628	.985	.948	.865	.684	.500
	$1-\beta_{SS}$	.776	.655	.545	.415	.323	.874	.774	.664	.478	.340
	$1-\beta_{NSS}$	.766	.650	.533	.400	.309	.887	.762	.626	.378	.215
≠ n's	α	.067	.037	.020	.008	.006	.076	.033	.015	.005	.002
≠ σ_e^2 (stepwise)	$1-\beta_{SL}$	.938	.887	.817	.709	.625	.970	.932	.878	.768	.649
	$1-\beta_{NSL}$	.938	.880	.805	.714	.640	.981	.941	.856	.670	.497
positively related	$1-\beta_{SS}$	.775	.653	.546	.386	.286	.883	.794	.683	.478	.332
	$1-\beta_{NSS}$	.766	.664	.561	.424	.305	.886	.764	.620	.418	.250
≠ n's	α	.099	.055	.035	.017	.011	.122	.048	.017	.004	.003
≠ σ_e^2 (stepwise)	$1-\beta_{SL}$	.968	.935	.899	.806	.728	.992	.973	.931	.833	.732
	$1-\beta_{NSL}$	.830	.741	.633	.494	.418	.866	.632	.329	.096	.041
negatively related	$1-\beta_{SS}$	.816	.728	.619	.488	.384	.907	.809	.704	.534	.386
	$1-\beta_{NSS}$	.590	.449	.339	.230	.167	.591	.345	.178	.051	.017
≠ n's	α	.078	.034	.018	.007	.005	.090	.036	.013	.005	.001
≠ σ_e^2 (SSSL)	$1-\beta_{SL}$	.955	.914	.850	.747	.663	.975	.940	.886	.774	.640
	$1-\beta_{NSL}$	.958	.911	.848	.742	.671	.981	.954	.883	.722	.558
positively related	$1-\beta_{SS}$	.796	.675	.561	.438	.333	.881	.795	.675	.496	.349
	$1-\beta_{NSS}$	.787	.662	.543	.408	.310	.890	.796	.646	.394	.234
≠ n's	α	.111	.056	.025	.007	.005	.101	.051	.021	.004	.002
≠ σ_e^2 (SSSL)	$1-\beta_{SL}$	.968	.930	.878	.798	.730	.988	.967	.926	.820	.703
	$1-\beta_{NSL}$	.875	.781	.701	.571	.464	.890	.671	.394	.118	.036
negatively related	$1-\beta_{SS}$	.834	.764	.681	.538	.454	.907	.829	.713	.541	.413
	$1-\beta_{NSS}$	.639	.523	.420	.301	.234	.641	.402	.205	.058	.020
≠ n's	α	.074	.042	.020	.010	.006	.083	.033	.015	.007	.002
≠ σ_e^2 (SLLL)	$1-\beta_{SL}$	.941	.900	.827	.732	.655	.978	.949	.909	.790	.652
	$1-\beta_{NSL}$	.942	.889	.818	.706	.622	.983	.940	.851	.677	.513
positively related	$1-\beta_{SS}$	.770	.661	.549	.408	.298	.887	.783	.673	.479	.357
	$1-\beta_{NSS}$	.757	.658	.527	.386	.284	.904	.775	.616	.362	.214
≠ n's	α	.088	.047	.021	.015	.010	.095	.051	.024	.007	.003
≠ σ_e^2 (SLLL)	$1-\beta_{SL}$	.965	.932	.878	.777	.702	.992	.971	.935	.831	.726
	$1-\beta_{NSL}$	.812	.708	.608	.493	.387	.836	.624	.316	.100	.041
negatively related	$1-\beta_{SS}$	.845	.740	.620	.479	.375	.928	.837	.732	.552	.401
	$1-\beta_{NSS}$	.587	.471	.366	.254	.192	.633	.386	.211	.068	.021

TABLE 9

Normal Population-Unequal Sample Sizes, (6, 6, 6, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and The Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.113	.061	.032	.010	.009	.095	.053	.020	.006	.001
	$1-\beta_{SL}$	.948	.900	.824	.679	.575	.936	.853	.743	.540	.399
$= \sigma^2$	$1-\beta_{NSL}$	.940	.873	.790	.680	.564	.921	.835	.736	.546	.388
	$1-\beta_{SS}$	.724	.603	.486	.327	.246	.696	.534	.387	.231	.143
	$1-\beta_{NSS}$	.739	.617	.479	.327	.240	.720	.552	.402	.232	.140
$\neq$ n's	α	.075	.035	.020	.012	.005	.071	.036	.015	.003	.001
$\neq \sigma^2$ (stepwise)	$1-\beta_{SL}$	.936	.857	.745	.596	.486	.928	.842	.722	.524	.374
	$1-\beta_{NSL}$	.895	.800	.701	.566	.482	.840	.737	.603	.448	.320
positively related	$1-\beta_{SS}$	.696	.565	.417	.287	.196	.706	.541	.379	.225	.129
	$1-\beta_{NSS}$	.665	.521	.395	.258	.190	.607	.455	.308	.168	.107
$\neq$ n's	α	.109	.052	.031	.010	.003	.102	.045	.018	.003	.000
$\neq \sigma^2$ (stepwise)	$1-\beta_{SL}$	.960	.904	.812	.683	.588	.939	.847	.727	.550	.417
	$1-\beta_{NSL}$	.853	.769	.665	.525	.414	.797	.668	.518	.303	.163
negatively related	$1-\beta_{SS}$	.762	.646	.518	.366	.268	.739	.583	.443	.240	.140
	$1-\beta_{NSS}$	.606	.483	.361	.237	.175	.534	.389	.248	.116	.062
$\neq$ n's	α	.057	.033	.021	.009	.006	.069	.032	.011	.005	.002
$\neq \sigma^2$ (SSSL)	$1-\beta_{SL}$	.852	.773	.632	.482	.396	.875	.771	.623	.442	.315
	$1-\beta_{NSL}$	.831	.728	.636	.512	.427	.753	.631	.508	.341	.226
positively related	$1-\beta_{SS}$	.559	.444	.327	.215	.156	.607	.430	.304	.167	.106
	$1-\beta_{NSS}$	.569	.445	.337	.235	.175	.506	.364	.257	.153	.099
$\neq$ n's	α	.103	.064	.030	.012	.007	.105	.058	.022	.005	.003
$\neq \sigma^2$ (SSSL)	$1-\beta_{SL}$	.940	.873	.812	.693	.589	.916	.836	.732	.556	.421
	$1-\beta_{NSL}$	.832	.758	.652	.500	.424	.771	.642	.495	.300	.156
negatively related	$1-\beta_{SS}$	.753	.626	.501	.367	.271	.707	.547	.402	.249	.147
	$1-\beta_{NSS}$	.638	.496	.382	.253	.173	.552	.394	.251	.121	.066
$\neq$ n's	α	.102	.047	.021	.009	.006	.087	.032	.019	.005	.001
$\neq \sigma^2$ (SLLL)	$1-\beta_{SL}$	.942	.894	.803	.641	.527	.961	.900	.787	.568	.378
	$1-\beta_{NSL}$	.935	.872	.795	.644	.532	.917	.843	.713	.507	.351
positively related	$1-\beta_{SS}$	.709	.576	.432	.293	.206	.744	.596	.427	.238	.135
	$1-\beta_{NSS}$	.693	.557	.432	.301	.219	.669	.517	.352	.201	.125
$\neq$ n's	α	.101	.048	.028	.011	.007	.089	.036	.012	.007	.003
$\neq \sigma^2$ (SLLL)	$1-\beta_{SL}$	.945	.882	.831	.705	.591	.923	.854	.748	.557	.410
	$1-\beta_{NSL}$	.875	.795	.693	.540	.439	.844	.698	.540	.335	.181
negatively related	$1-\beta_{SS}$	.746	.622	.498	.363	.256	.729	.580	.420	.233	.135
	$1-\beta_{NSS}$	.641	.490	.363	.237	.172	.577	.400	.251	.137	.065

TABLE 10

Exponential Population-Unequal Sample Sizes, (6, 6, 6, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power (1-β). Power is based on four cases of mean differences (SL-stepwise, large;NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases (= σ_e^2; ≠ σ_e^2, stepwise; ≠ σ_e^2, SSSL, approximately in the ratio of 1:1:1:4; and ≠ σ_e^2, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
≠ n's = σ^2	α	.088	.035	.019	.006	.005	.100	.042	.021	.006	.002
	$1-\beta_{SL}$	.923	.875	.813	.710	.635	.961	.929	.878	.751	.642
	$1-\beta_{NSL}$	.942	.893	.820	.701	.603	.980	.943	.871	.705	.563
	$1-\beta_{SS}$	.757	.646	.557	.393	.307	.856	.771	.655	.481	.350
	$1-\beta_{NSS}$	.763	.651	.551	.392	.316	.908	.806	.648	.453	.319
≠ n's	α	.071	.036	.017	.006	.002	.073	.031	.013	.005	.004
	$1-\beta_{SL}$	.956	.896	.836	.747	.668	.984	.951	.887	.784	.672
≠ σ^2 (stepwise)	$1-\beta_{NSL}$	.936	.883	.805	.694	.600	.979	.947	.888	.727	.584
	$1-\beta_{SS}$	.756	.648	.537	.416	.314	.861	.765	.647	.491	.373
positively related	$1-\beta_{NSS}$	.775	.643	.539	.384	.297	.892	.785	.650	.445	.294
≠ n's	α	.096	.054	.031	.015	.005	.098	.039	.017	.006	.002
≠ σ^2 (stepwise)	$1-\beta_{SL}$	.955	.915	.864	.770	.672	.986	.963	.903	.802	.679
	$1-\beta_{NSL}$	.892	.813	.717	.594	.503	.943	.855	.675	.410	.217
negatively related	$1-\beta_{SS}$	.806	.714	.597	.458	.375	.896	.819	.701	.509	.359
	$1-\beta_{NSS}$	.678	.546	.433	.289	.210	.777	.608	.398	.187	.067
≠ n's	α	.081	.040	.020	.012	.006	.085	.036	.012	.005	.000
≠ σ^2 (SSSL)	$1-\beta_{SL}$	.950	.898	.834	.740	.650	.978	.945	.879	.770	.657
	$1-\beta_{NSL}$	.957	.905	.833	.724	.624	.990	.950	.890	.748	.596
positively related	$1-\beta_{SS}$	.778	.677	.572	.424	.316	.873	.786	.658	.504	.386
	$1-\beta_{NSS}$	.793	.661	.546	.398	.316	.886	.796	.677	.466	.310
≠ n's	α	.097	.052	.024	.011	.005	.104	.050	.025	.011	.003
≠ σ^2 (SSSL)	$1-\beta_{SL}$	.946	.907	.844	.750	.683	.982	.949	.892	.781	.671
	$1-\beta_{NSL}$	.875	.800	.707	.577	.497	.936	.841	.673	.387	.182
negatively related	$1-\beta_{SS}$	.800	.717	.598	.433	.337	.897	.805	.669	.487	.334
	$1-\beta_{NSS}$	.671	.537	.424	.309	.232	.775	.599	.394	.168	.070
≠ n's	α	.077	.048	.024	.009	.005	.089	.037	.012	.003	.000
≠ σ^2 (SLLL)	$1-\beta_{SL}$	.933	.890	.820	.735	.659	.975	.930	.872	.771	.673
	$1-\beta_{NSL}$	.928	.862	.792	.686	.601	.972	.931	.870	.719	.571
positively related	$1-\beta_{SS}$	.753	.641	.521	.387	.292	.862	.757	.650	.482	.352
	$1-\beta_{NSS}$	.782	.657	.547	.421	.315	.897	.808	.671	.461	.320
≠ n's	α	.102	.045	.028	.016	.005	.106	.048	.023	.007	.002
≠ σ^2 (SLLL)	$1-\beta_{SL}$	.950	.907	.848	.755	.678	.985	.957	.908	.782	.654
	$1-\beta_{NSL}$	.884	.797	.714	.589	.503	.945	.859	.678	.408	.207
negatively related	$1-\beta_{SS}$	.831	.724	.599	.451	.348	.911	.803	.673	.501	.368
	$1-\beta_{NSS}$	.687	.578	.479	.348	.241	.790	.634	.449	.190	.090

TABLE 11

Normal Population-Unequal Sample Sizes

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power (1-β). Power is based on four cases of mean differences (SL-stepwise, large;NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed for stepwise inequality of variances which are positively and negatively related to sample sizes.

Sample Sizes = 11,15,19,23

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
≠ n's	α	.074	.039	.017	.005	.002	.079	.037	.018	.005	.002
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.915	.819	.715	.586	.484	.917	.844	.755	.598	.487
	$1-\beta_{NSL}$	.877	.803	.706	.570	.487	.849	.743	.623	.487	.384
positively related	$1-\beta_{SS}$	.666	.535	.424	.296	.215	.692	.545	.438	.294	.206
	$1-\beta_{NSS}$	.652	.529	.409	.278	.211	.617	.481	.332	.233	.158
≠ n's	α	.123	.064	.033	.014	.005	.107	.044	.027	.007	.003
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.941	.895	.828	.730	.639	.925	.865	.788	.660	.542
	$1-\beta_{NSL}$	.774	.676	.568	.467	.387	.706	.579	.468	.321	.239
negatively related	$1-\beta_{SS}$	.762	.657	.545	.411	.310	.720	.604	.478	.341	.237
	$1-\beta_{NSS}$	.530	.410	.322	.214	.175	.454	.329	.223	.135	.093

Sample Sizes = 11,19,19,19

≠ n's	α	.099	.041	.023	.008	.003	.093	.047	.024	.006	.001
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.920	.856	.776	.663	.572	.910	.839	.744	.608	.495
	$1-\beta_{NSL}$	.920	.849	.771	.658	.571	.895	.813	.699	.571	.465
positively related	$1-\beta_{SS}$	.750	.615	.489	.340	.261	.724	.584	.455	.298	.217
	$1-\beta_{NSS}$	.723	.587	.454	.329	.257	.683	.542	.397	.263	.191
≠ n's	α	.115	.061	.035	.014	.007	.113	.060	.029	.008	.005
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.931	.868	.794	.688	.608	.915	.842	.764	.632	.530
	$1-\beta_{NSL}$	.817	.721	.611	.466	.374	.773	.652	.528	.365	.261
negatively related	$1-\beta_{SS}$	.747	.621	.514	.361	.289	.719	.590	.464	.311	.237
	$1-\beta_{NSS}$	.579	.438	.337	.212	.150	.529	.385	.269	.153	.100

TABLE 12

Exponential Population-Unequal Sample Sizes

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed for stepwise inequality of variances which are positively and negatively related to sample sizes.

Sample Sizes = 11,15,19,23

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.079	.036	.012	.007	.005	.069	.033	.016	.010	.007
	$1-\beta_{SL}$	.983	.964	.913	.828	.731	.999	.993	.985	.964	.932
$\neq \sigma^2$ (stepwise)	$1-\beta_{NSL}$	.937	.889	.825	.732	.663	.994	.979	.957	.911	.863
	$1-\beta_{SS}$	.865	.739	.619	.457	.355	.961	.914	.848	.748	.658
positively related	$1-\beta_{NSS}$	.776	.667	.547	.406	.306	.928	.858	.750	.619	.504
$\neq$ n's	α	.122	.061	.033	.018	.010	.119	.061	.031	.013	.006
	$1-\beta_{SL}$	.982	.954	.918	.844	.781	.997	.992	.984	.963	.938
$\neq \sigma^2$ (stepwise)	$1-\beta_{NSL}$	.843	.745	.641	.499	.414	.970	.915	.823	.632	.491
	$1-\beta_{SS}$	.844	.762	.668	.531	.446	.965	.918	.840	.721	.620
negatively related	$1-\beta_{NSS}$	.611	.501	.402	.275	.224	.797	.650	.500	.302	.184

Sample Sizes = 11,19,19,19

$\neq$ n's	α	.087	.049	.024	.009	.003	.092	.047	.017	.008	.004
	$1-\beta_{SL}$	.954	.908	.845	.747	.668	.993	.984	.964	.930	.886
$\neq \sigma^2$ (stepwise)	$1-\beta_{NSL}$	.939	.881	.819	.719	.636	.991	.984	.963	.921	.868
	$1-\beta_{SS}$	.777	.667	.563	.431	.334	.934	.894	.812	.708	.598
positively related	$1-\beta_{NSS}$	.753	.657	.561	.412	.342	.951	.883	.808	.676	.552
$\neq$ n's	α	.086	.037	.016	.007	.001	.103	.049	.025	.006	.002
	$1-\beta_{SL}$	.930	.882	.814	.721	.651	.993	.982	.961	.906	.858
$\neq \sigma^2$ (stepwise)	$1-\beta_{NSL}$	.824	.716	.616	.487	.407	.973	.919	.832	.659	.513
	$1-\beta_{SS}$	.786	.677	.577	.444	.360	.935	.883	.814	.689	.583
negatively related	$1-\beta_{NSS}$	.604	.474	.362	.255	.189	.828	.692	.530	.342	.219

TABLE 13

Schematic of Conditions Investigated ("X" indicates inclusion of the case. Each case was investigated for a normal population and an exponential population using both the ANOVA F-test and the Kruskal-Wallis Test).

			$\neq n_j$'s					
		[7,7,7,7]	[4,6,8,10]	[4,8,8,8]	[6,6,6,10]	[11,15,19,23]	[11,19,19,19]	
$= \sigma_e^2$	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X			
$\neq \sigma_e^2$ stepwise	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X	X	X	pos. related n_j's and σ_e^2's
			X	X	X	X	X	neg. related n_j's and σ_e^2's
$\neq \sigma_e^2$ SLLL	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X			pos. related n_j's and σ_e^2's
			X	X	X			neg. related n_j's and σ_e^2's
$\neq \sigma_e^2$ SSSL	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X			pos. related n_j's and σ_e^2's
			X	X	X			neg. related n_j's and σ_e^2's

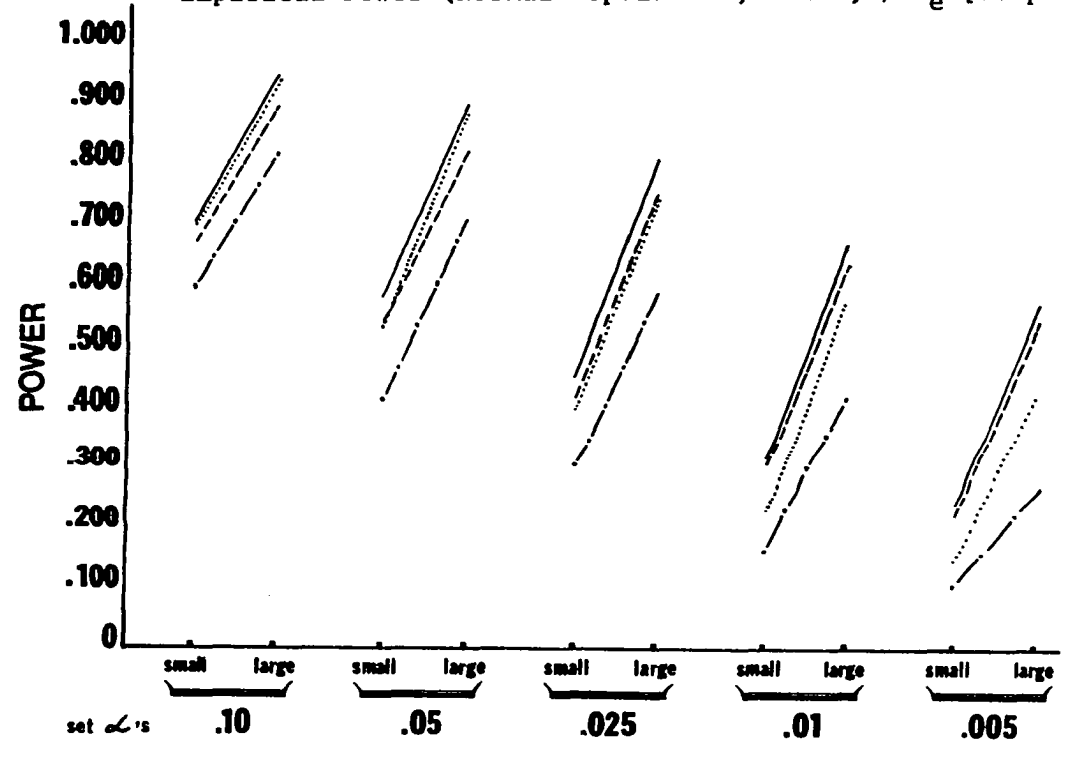
n_j - sample size per treatment level

SL - a stepwise difference such that power was approximately .86 for α=.05 for the normal distribution
NSL - a nonstepwise difference (1:1:1:4) such that power was approximately .86 for α=.05 for the normal distribution
SS - a stepwise difference such that power was approximately .60 for α=.05 for the normal distribution
NSS - a nonstepwise difference (1:1:1:4) such that power was approximately .60 for α=.05 for the normal distribution

Stepwise - variances in the ratio of 1:2:3:4
SLLL - variances in the ratio of 1:4:4:4
SLLL - variances in the ratio of 1:1:1:4

FIGURE 1

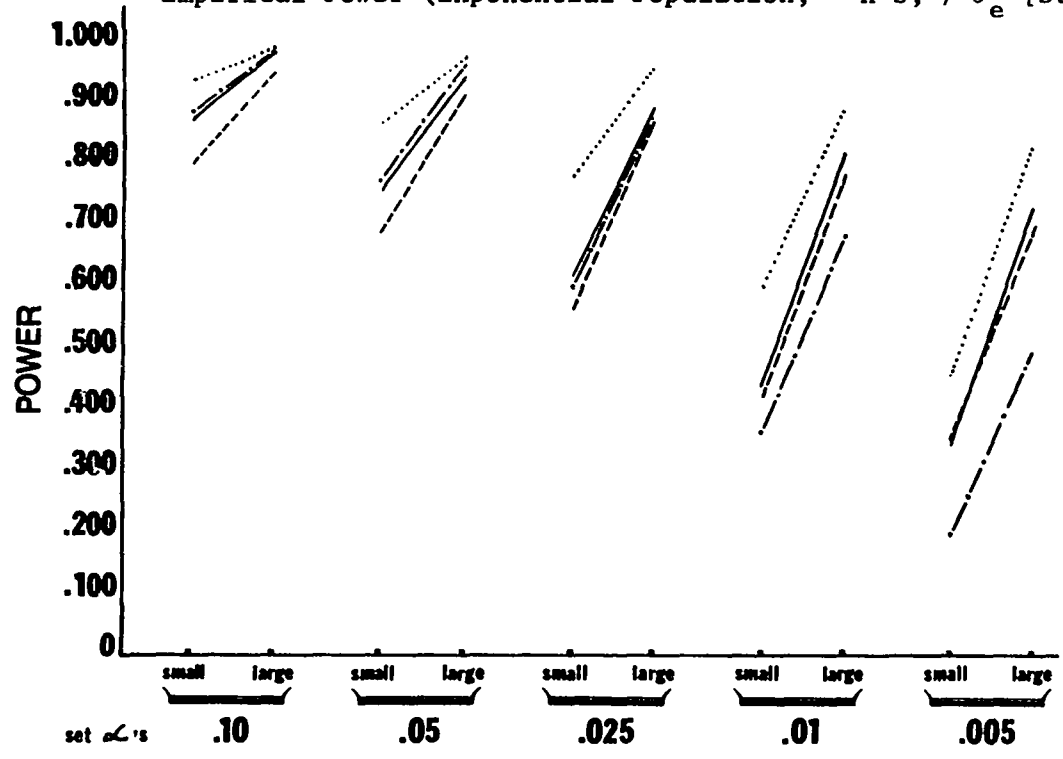
Empirical Power (Normal Population; = n's, ≠ σ_e^2 [stepwise])



LEGEND
——— Power for the ANOVA F-test based on stepwise mean differences
– – – – – Power for the ANOVA F-test based on nonstepwise mean differences
..... Power for the Kruskal-Wallis test based on stepwise mean differences
•—•—• Power for the Kruskal-Wallis test based on nonstepwise mean differences
small Small mean differences (power = .60 for a normal population at .05 level)
large Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 2

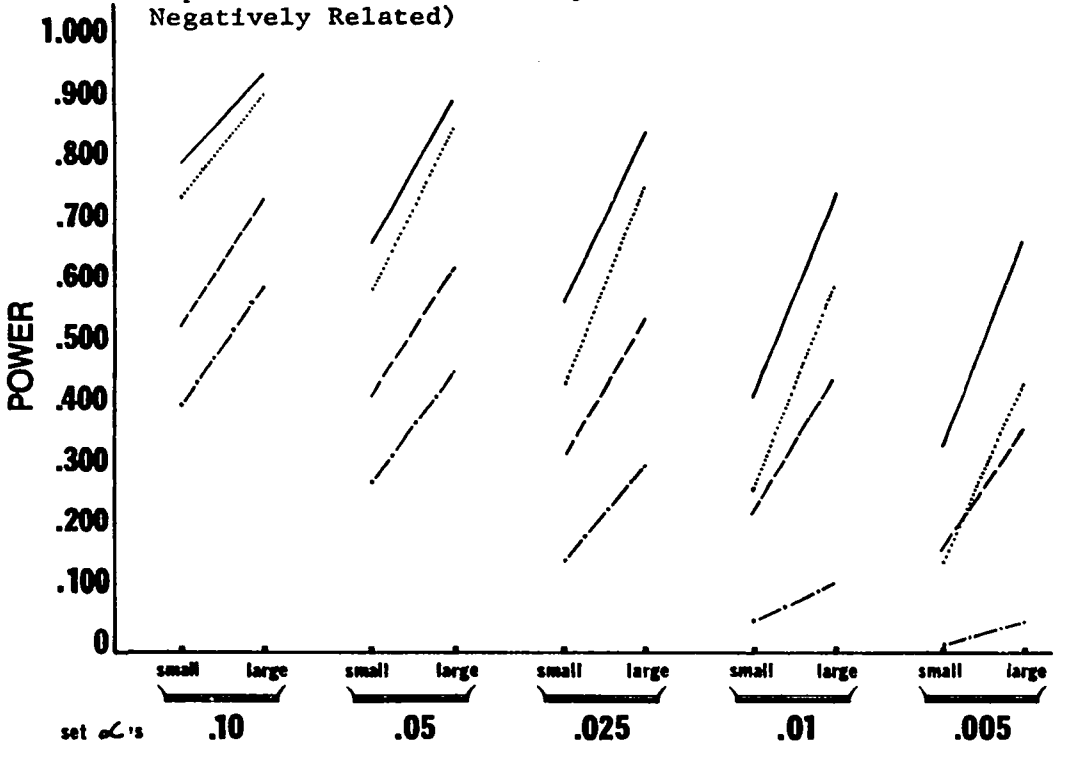
Empirical Power (Exponential Population; = n's, $\neq \sigma_e^2$ [stepwise])



LEGEND
——— Power for the ANOVA F-test based on stepwise mean differences
- - - - Power for the ANOVA F-test based on nonstepwise mean differences
····· Power for the Kruskal-Wallis test based on stepwise mean differences
•—•—• Power for the Kruskal-Wallis test based on nonstepwise mean differences
small Small mean differences (power = .60 for a normal population at .05 level)
large Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 3

Empirical Power (Normal Population; ≠ n's [4,6,8,10], ≠ σ_e^2 [stepwise]; Negatively Related)

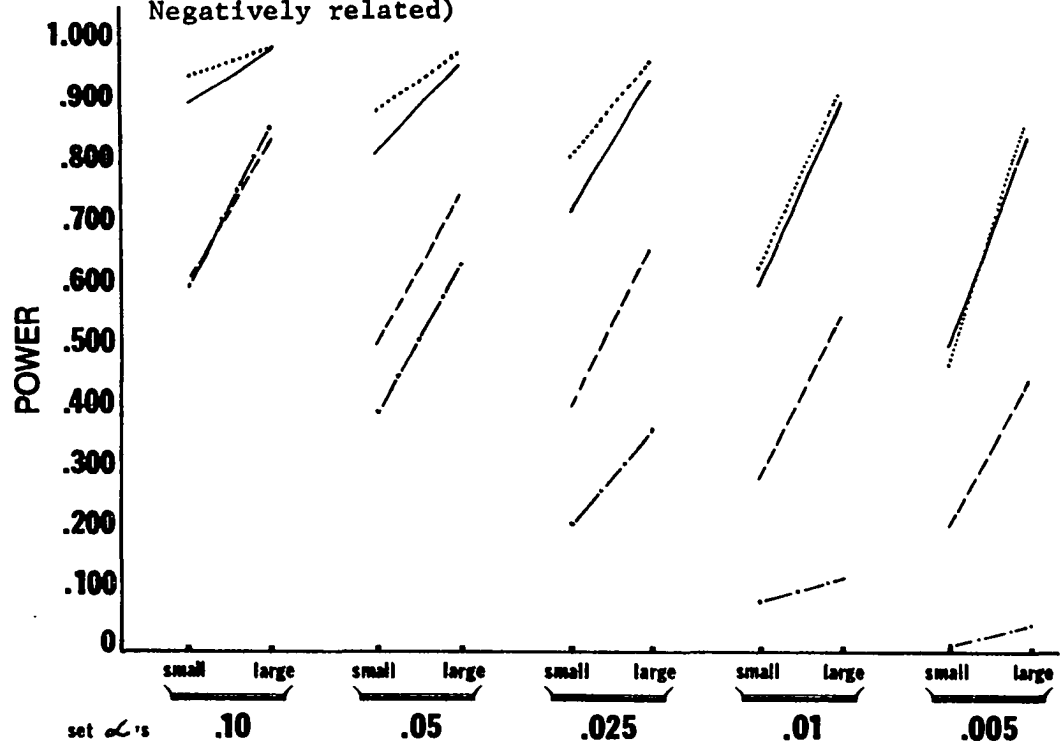


LEGEND

———	Power for the ANOVA F-test based on stepwise mean differences
- - - -	Power for the ANOVA F-test based on nonstepwise mean differences
.....	Power for the Kruskal-Wallis test based on stepwise mean differences
·—·—·	Power for the Kruskal-Wallis test based on nonstepwise mean differences
small	Small mean differences (power = .60 for a normal population at .05 level)
large	Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 4

Empirical Power (Exponential Population; ≠ n's [4,6,8,10], ≠ σ_e^2 [stepwise]; Negatively related)

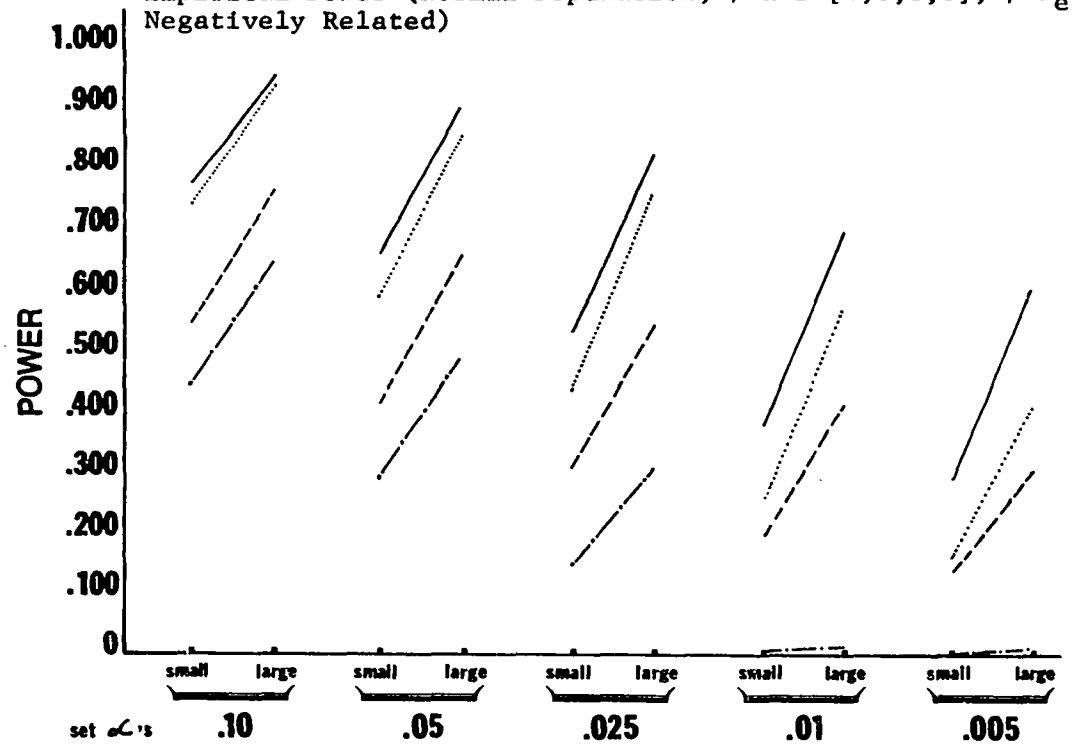


LEGEND

———	Power for the ANOVA F-test based on stepwise mean differences
- - - -	Power for the ANOVA F-test based on nonstepwise mean differences
·····	Power for the Kruskal-Wallis test based on stepwise mean differences
-·-·-	Power for the Kruskal-Wallis test based on nonstepwise mean differences
small	Small mean differences (power = .60 for a normal population at .05 level)
large	Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 5

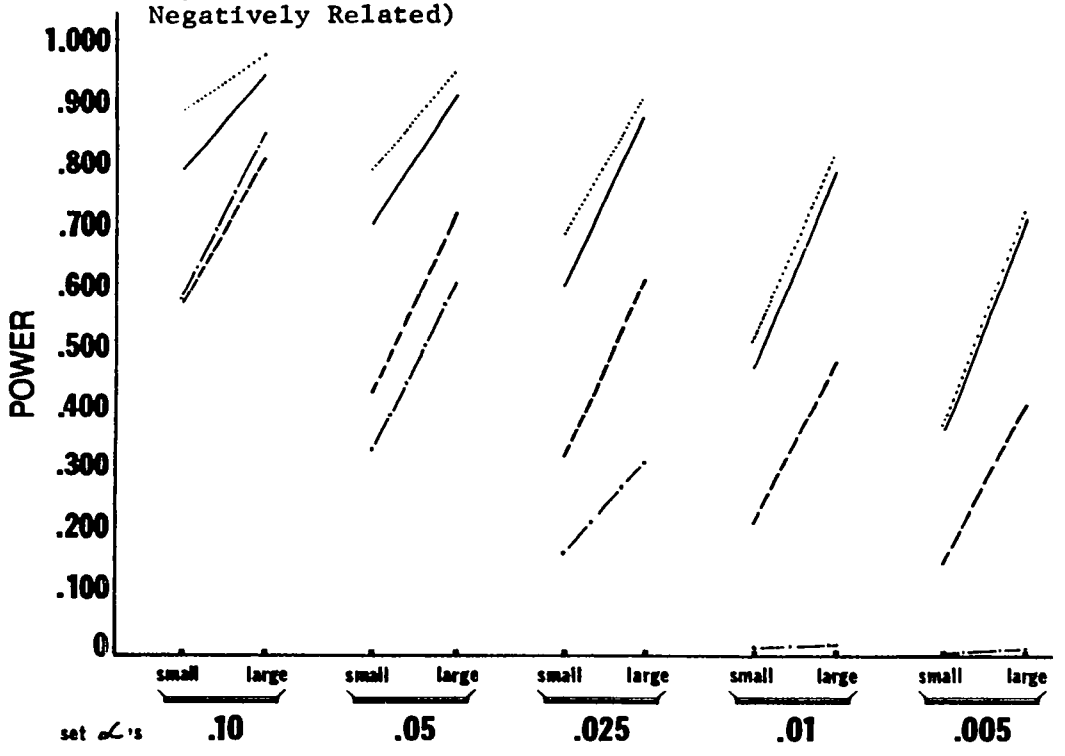
Empirical Power (Normal Population; ≠ n's [4,8,8,8], ≠ σ_e^2 [stepwise]; Negatively Related)



LEGEND
——— Power for the ANOVA F-test based on stepwise mean differences
– – – – Power for the ANOVA F-test based on nonstepwise mean differences
..... Power for the Kruskal-Wallis test based on stepwise mean differences
•—•—• Power for the Kruskal-Wallis test based on nonstepwise mean differences
small Small mean differences (power = .60 for a normal population at .05 level)
large Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 6

Empirical Power (Exponential Population; ≠ n's [4,8,8,8], ≠ σ_e^2 [stepwise]; Negatively Related)



LEGEND

———	Power for the ANOVA F-test based on stepwise mean differences
- - - -	Power for the ANOVA F-test based on nonstepwise mean differences
.....	Power for the Kruskal-Wallis test based on stepwise mean differences
·—·—·	Power for the Kruskal-Wallis test based on nonstepwise mean differences
small	Small mean differences (power = .60 for a normal population at .05 level)
large	Large mean differences (power = .86 for a normal population at .05 level)